# Machine Learning in Particle Physics

Tilman Plehn

Universität Heidelberg

University of Sao Paulo, November 2025

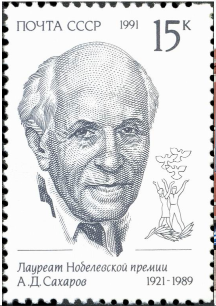
# LHC & simulations

## Classic motivation

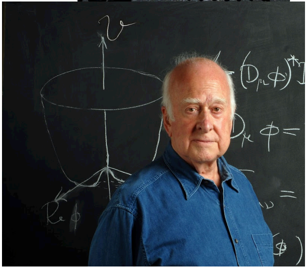
- dark matter?
- matter vs antimatter?
- origin of Higgs boson?

## Strengths

- fundamental questions
- huge, complex data set
- first-principle, precision simulations







# LHC & simulations

## Classic motivation

- dark matter?
- matter vs antimatter?
- origin of Higgs boson?

## Strengths

- fundamental questions
- huge, complex data set
- first-principle, precision simulations

## First-principle simulations

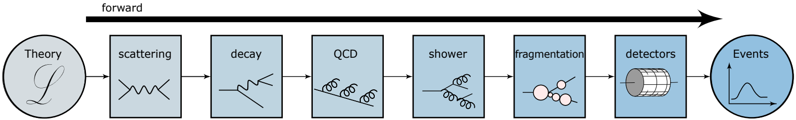
- start with Lagrangian
- calculate scattering using QFT
- simulate collisions
- simulate detectors

→ LHC events in virtual worlds

## Searches and measurements

- compare simulations and data
- infer underlying theory [SM or BSM]
- publish data to re-interpret

→ Understand LHC data systematically



# Neural networks

## Similar to fit

- approximate $f_\theta(x) \approx f(x)$
- $x$ low-D phase space
- $f_\theta$ numerical function

$\to$ $\theta$ data representation

## Probabilities over phase space

- regression $\quad x \to A_\theta(x)$
- classification $\quad x \to p_\theta(x)$ [likelihood ratio]
- generation $\quad r \sim \mathcal{N} \to x \sim p_\theta(x)$
- conditional generation $\quad r \sim \mathcal{N} \to x \sim p_\theta(x|y)$

## LHC representations

- accuracy
- precision
- structure

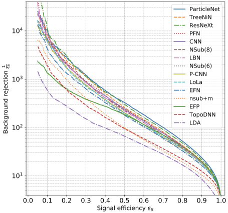
$\to$ Physics-specific ML

# ML in experiment

## Top tagging [classification, 2016-today]

- ‘hello world’ of LHC-ML
- end of QCD-taggers
- ever-improving [Huilin Qu]

→ Driving NN-architectures

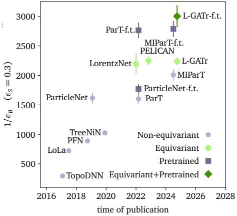



**SciPost Physics** **Submission**

### The Machine Learning Landscape of Top Taggers

G. Kasieczka (ed)[1], T. Plehn (ed)[2], A. Butter[2], K. Cranmer[3], D. Debnath[4], B. M. Dillon[5], M. Fairbairn[6], D. A. Faroughy[5], W. Fedorko[7], C. Gay[7], L. Gouskos[8], J. F. Kamenik[5,9], P. T. Komiske[10], S. Leiss[1], A. Lister[7], S. Macaluso[3,4], E. M. Metodiev[10], L. Moore[11], B. Nachman,[12,13], K. Nordström[14,15], J. Pearkes[7], H. Qu[8], Y. Rath[16], M. Rieger[16], D. Shih[4], J. M. Thompson[2], and S. Varma[6]

**1** Institut für Experimentalphysik, Universität Hamburg, Germany
**2** Institut für Theoretische Physik, Universität Heidelberg, Germany
**3** Center for Cosmology and Particle Physics and Center for Data Science, NYU, USA
**4** NHECT, Dept. of Physics and Astronomy, Rutgers, The State University of NJ, USA
**5** Jozef Stefan Institute, Ljubljana, Slovenia
**6** Theoretical Particle Physics and Cosmology, King’s College London, United Kingdom
**7** Department of Physics and Astronomy, The University of British Columbia, Canada
**8** Department of Physics, University of California, Santa Barbara, USA

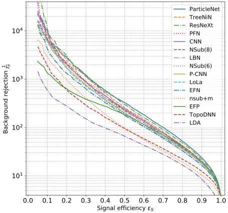


# ML in experiment

## Top tagging [classification, 2016-today]

- 'hello world' of LHC-ML
- end of QCD-taggers
- ever-improving [Huilin Qu]

→ Driving NN-architectures




SciPost Physics | Submission

**The Machine Learning Landscape of Top Taggers**

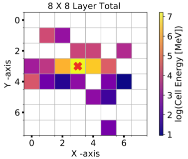
G. Kasieczka (ed)[1], T. Plehn (ed)[2], A. Butter[2], K. Cranmer[3], D. Debnath[4], B. M. Dillon[5], M. Fairbairn[6], D. A. Faroughy[5], W. Fedorko[7], C. Gay[7], L. Gouskos[8], J. F. Kamenik[5,9], P. T. Komiske[10], S. Leiss[1], A. Lister[7], S. Macaluso[3,4], E. M. Metodiev[10], L. Moore[11], B. Nachman,[12,13], K. Nordström[14,15], J. Pearkes[7], H. Qu[8], Y. Rath[16], M. Rieger[16], D. Shih[4], J. M. Thompson[2], and S. Varma[6]

1 Institut für Experimentalphysik, Universität Hamburg, Germany
2 Institut für Theoretische Physik, Universität Heidelberg, Germany
3 Center for Cosmology and Particle Physics and Center for Data Science, NYU, USA
4 NHECT, Dept. of Physics and Astronomy, Rutgers, The State University of NJ, USA
5 Jozef Stefan Institute, Ljubljana, Slovenia
6 Theoretical Particle Physics and Cosmology, King's College London, United Kingdom
7 Department of Physics and Astronomy, The University of British Columbia, Canada
8 Department of Physics, University of California, Santa Barbara, USA
9 Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia
10 Center for Theoretical Physics, MIT, Cambridge, USA
11 CP3, Universitéxx Catholique de Louvain, Louvain-la-Neuve, Belgium
12 Physics Division, Lawrence Berkeley National Laboratory, Berkeley, USA
13 Simons Inst. for the Theory of Computing, University of California, Berkeley, USA
14 National Institute for Subatomic Physics (NIKHEF), Amsterdam, Netherlands
15 LPTHE, CNRS & Sorbonne Université, Paris, France
16 III. Physics Institute A, RWTH Aachen University, Germany

## Particle flow [2020-today]

- basis of jet analyses
- combining detectors with different resolution

→ Optimality the key



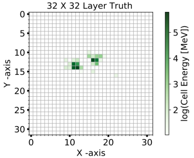

**Towards a Computer Vision Particle Flow** *

Francesco Armando Di Bello[a,3], Sanmay Ganguly[b,1], Eilam Gross[1], Marumi Kado[3,4], Michael Pitt[2], Lorenzo Santi [3], Jonathan Shlomi[1]

[1]Weizmann Institute of Science, Rehovot 76100, Israel
[2]CERN, CH 1211, Geneva 23, Switzerland
[3]Università di Roma Sapienza, Piazza Aldo Moro, 2, 00185 Roma, Italy e INFN, Italy
[4]Université Paris-Saclay, CNRS/IN2P3, IJCLab, 91405, Orsay, France

**Progress towards an improved particle flow algorithm at CMS with machine learning**

Farouk Mokhtar[1], Joosep Pata[2], Javier Duarte[1], Eric Wulff[3], Maurizio Pierini[3] and Jean-Roch Vlimant[4]
(on behalf of the CMS Collaboration)
[1]University of California San Diego, La Jolla, CA 92093, USA
[2]NICPB, Rävala pst 10, 10143 Tallinn, Estonia
[3]European Organization for Nuclear Research (CERN), CH 1211, Geneva 23, Switzerland
[4]California Institute of Technology, Pasadena, CA 91125, USA
E-mail: fmokhtar@ucsd.edu, joosep.pata@cern.ch, jduarte@ucsd.edu

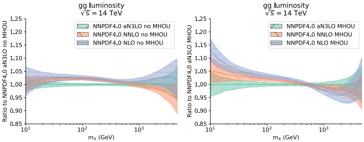


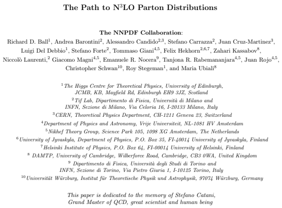
# ML in phenomenology

## Parton densities [NNPDF, 2002-today]

- LHC-ML classic
- pdfs with uncertainties and without bias

→ Driving precision



The Path to N$^3$LO Parton Distributions

The NNPDF Collaboration:

Richard D. Ball[1], Andrea Barontini[2], Alessandro Candido[2,3], Stefano Carrazza[2], Juan Cruz-Martinez[3], Luigi Del Debbio[1], Stefano Forte[2], Tommaso Giani[4,5], Felix Hekhorn[2,6,7], Zahari Kassabov[8], Niccolò Laurenti,[2] Giacomo Magni[4,5], Emanuele R. Nocera[9], Tanjona R. Rabemananjara[4,5], Juan Rojo[4,5], Christopher Schwan[10], Roy Stegeman[1], and Maria Ubiali[8]

[1] The Higgs Centre for Theoretical Physics, University of Edinburgh, JCMB, KB, Mayfield Rd, Edinburgh EH9 3JZ, Scotland
[2] Tif Lab, Dipartimento di Fisica, Università di Milano and INFN, Sezione di Milano, Via Celoria 16, I-20133 Milano, Italy
[3] CERN, Theoretical Physics Department, CH-1211 Geneva 23, Switzerland
[4] Department of Physics and Astronomy, Vrije Universiteit, NL-1081 HV Amsterdam
[5] Nikhef Theory Group, Science Park 105, 1098 XG Amsterdam, The Netherlands
[6] University of Jyvaskyla, Department of Physics, P.O. Box 35, FI-40014 University of Jyvaskyla, Finland
[7] Helsinki Institute of Physics, P.O. Box 64, FI-00014 University of Helsinki, Finland
[8] DAMTP, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, United Kingdom
[9] Dipartimento di Fisica, Università degli Studi di Torino and INFN, Sezione di Torino, Via Pietro Giuria 1, I-10125 Torino, Italy
[10] Universität Würzburg, Institut für Theoretische Physik und Astrophysik, 97074 Würzburg, Germany

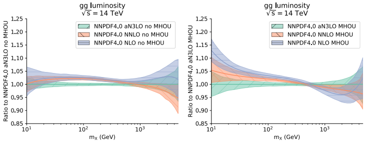
*This paper is dedicated to the memory of Stefano Catani, Grand Master of QCD, great scientist and human being*

# ML in phenomenology

## Parton densities [NNPDF, 2002-today]

- LHC-ML classic
- pdfs with uncertainties and without bias

→ Driving precision




**The Path to N$^3$LO Parton Distributions**

**The NNPDF Collaboration**:
Richard D. Ball[1], Andrea Barontini[2], Alessandro Candido[2,3], Stefano Carrazza[2], Juan Cruz-Martinez[3], Luigi Del Debbio[1], Stefano Forte[2], Tommaso Giani[4,5], Felix Hekhorn[2,6,7], Zahari Kassabov[8], Niccolò Laurenti,[2] Giacomo Magni[4,5], Emanuele R. Nocera[9], Tanjona R. Rabemananjara[4,5], Juan Rojo[4,5], Christopher Schwan[10], Roy Stegeman[1], and Maria Ubiali[8]

[1] *The Higgs Centre for Theoretical Physics, University of Edinburgh, JCMB, KB, Mayfield Rd, Edinburgh EH9 3JZ, Scotland*
[2] *Tif Lab, Dipartimento di Fisica, Università di Milano and INFN, Sezione di Milano, Via Celoria 16, I-20133 Milano, Italy*
[3] *CERN, Theoretical Physics Department, CH-1211 Geneva 23, Switzerland*
[4] *Department of Physics and Astronomy, Vrije Universiteit, NL-1081 HV Amsterdam*
[5] *Nikhef Theory Group, Science Park 105, 1098 XG Amsterdam, The Netherlands*
[6] *University of Jyvaskyla, Department of Physics, P.O. Box 35, FI-40014 University of Jyvaskyla, Finland*
[7] *Helsinki Institute of Physics, P.O. Box 64, FI-00014 University of Helsinki, Finland*
[8] *DAMTP, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, United Kingdom*
[9] *Dipartimento di Fisica, Università degli Studi di Torino and INFN, Sezione di Torino, Via Pietro Giuria 1, I-10125 Torino, Italy*
[10] *Universität Würzburg, Institut für Theoretische Physik und Astrophysik, 97074 Würzburg, Germany*

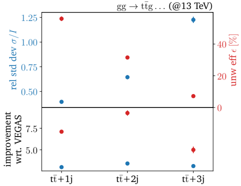
*This paper is dedicated to the memory of Stefano Catani, Grand Master of QCD, great scientist and human being*

## Ultra-fast simulations [Sherpa, MadNIS, MLHad]

- event generation modular
- better ML-modules

→ MadNIS → MadGraph7

| | | | |
|---|---|---|---|
| Triple-W | $u\bar{d} \to W^+W^+W^-$ | | |
| VBS | $uc \to W^+W^+\,ds$ | | |
| W+jets | $gg \to W^+d\bar{u}$ | $gg \to W^+d\bar{u}g$ | $gg \to W^+d\bar{u}gg$ |
| $t\bar{t}$+jets | $gg \to t\bar{t}+g$ | $gg \to t\bar{t}+gg$ | $gg \to t\bar{t}+ggg$ , |




SciPost Physics Submission

**The MadNIS Reloaded**

Theo Heimel[1], Nathan Huetsch[1], Fabio Maltoni[2,3], Olivier Mattelaer[2], Tilman Plehn[1], and Ramon Winterhalder[2]

**1** Institut für Theoretische Physik, Universität Heidelberg, Germany
**2** CP3, Université catholique de Louvain, Louvain-la-Neuve, Belgium
**3** Dipartimento di Fisica e Astronomia, Universitá di Bologna, Italy

December 17, 2024

**Abstract**

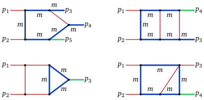
**In pursuit of precise and fast theory predictions for the LHC, we present an implementation of the MadNIS method in the MadGraph event generator. A series of improvements in MadNIS further enhance its efficiency and speed. We validate this implementation for realistic partonic processes and find significant gains from using modern machine learning in event generators.**

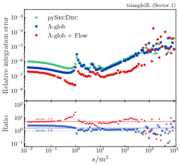
# ML in theory

## Optimizing integration paths [invertible networks]

- compute Feynman integrals
- learn optimal integration pat

→ To be implemented...




SciPost SciPost Phys. 12, 129 (2022)

### Targeting multi-loop integrals with neural networks

Ramon Winterhalder[1,2,3], Vitaly Magerya[4], Emilio Villa[4], Stephen P. Jones[5], Matthias Kerner[4,6], Anja Butter[1,2], Gudrun Heinrich[2,4] and Tilman Plehn[1,2]

1 Institut für Theoretische Physik, Universität Heidelberg, Germany
2 HEiKA - Heidelberg Karlsruhe Strategic Partnership, Heidelberg University, Karlsruhe Institute of Technology (KIT), Germany
3 Centre for Cosmology, Particle Physics and Phenomenology (CP3), Université catholique de Louvain, Belgium
4 Institut für Theoretische Physik, Karlsruher Institut für Technologie, Germany
5 Institute for Particle Physics Phenomenology, Durham University, UK
6 Institut für Astroteilchenphysik, Karlsruher Institut für Technologie, Germany

**Abstract**

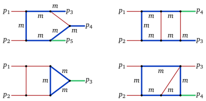
Numerical evaluations of Feynman integrals often proceed via a deformation of the integration contour into the complex plane. While valid contours are easy to construct, the numerical precision for a multi-loop integral can depend critically on the chosen contour. We present methods to optimize this contour using a combination of optimized, global complex shifts and a normalizing flow. They can lead to a significant gain in precision.

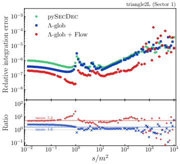
# ML in theory

## Optimizing integration paths [invertible networks]

- compute Feynman integrals
- learn optimal integration path

→ To be implemented...





SciPost Phys. 12, 129 (2022)

**Targeting multi-loop integrals with neural networks**

Ramon Winterhalder[1,2,3], Vitaly Magerya[4], Emilio Villa[4], Stephen P. Jones[5], Matthias Kerner[4,6], Anja Butter[1,2], Gudrun Heinrich[2,4] and Tilman Plehn[1,2]

1 Institut für Theoretische Physik, Universität Heidelberg, Germany
2 HEiKA - Heidelberg Karlsruhe Strategic Partnership, Heidelberg University, Karlsruhe Institute of Technology (KIT), Germany
3 Centre for Cosmology, Particle Physics and Phenomenology (CP3), Université catholique de Louvain, Belgium
4 Institut für Theoretische Physik, Karlsruher Institut für Technologie, Germany
5 Institute for Particle Physics Phenomenology, Durham University, UK
6 Institut für Astroteilchenphysik, Karlsruher Institut für Technologie, Germany

**Abstract**

Numerical evaluations of Feynman integrals often proceed via a deformation of the integration contour into the complex plane. While valid contours are easy to construct, the numerical precision for a multi-loop integral can depend critically on the chosen contour. We present methods to optimize this contour using a combination of optimized, global complex shifts and a normalizing flow. They can lead to a significant gain in precision.

## String landscape [reinforcement learning]

- searching for viable vacua
- high dimensions, unknown global structure

→ Islands of Standard Model?

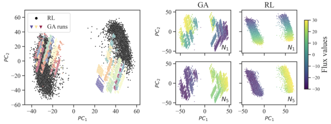
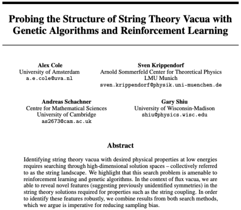



Figure 1: *Left:* Cluster structure in dimensionally reduced flux samples for RL and 25 GA runs (PCA on all samples of GA and RL). The colors indicate individual GA runs. *Right:* Dependence on flux (input) values ($N_3$ and $N_5$ respectively) in relation to principal components for a PCA fit of the individual output of GA and RL.


**Probing the Structure of String Theory Vacua with Genetic Algorithms and Reinforcement Learning**

Alex Cole
University of Amsterdam
a.e.cole@uva.nl

Sven Krippendorf
Arnold Sommerfeld Center for Theoretical Physics
LMU Munich
sven.krippendorf@physik.uni-muenchen.de

Andreas Schachner
Centre for Mathematical Sciences
University of Cambridge
as2673@cam.ac.uk

Gary Shiu
University of Wisconsin-Madison
shiu@physics.wisc.edu

**Abstract**

Identifying string theory vacua with desired physical properties at low energies requires searching through high-dimensional solution spaces – collectively referred to as the string landscape. We highlight that this search problem is amenable to reinforcement learning and genetic algorithms. In the context of flux vacua, we are able to reveal novel features (suggesting previously unidentified symmetries) in the string theory solutions required for properties such as the string coupling. In order to identify these features robustly, we combine results from both search methods, which we argue is imperative for reducing sampling bias.

# Training and uncertainties

## Learned scalar field $f_\theta(x) \approx f(x)$

- maximize parameter probability given $(f_j, \sigma_j)$

$$\theta = \text{argmax}\; p(\theta|x) = \text{argmax}\; \frac{p(x|\theta)\; p(\theta)}{p(x)}$$

$\rightarrow$ Gaussian likelihood loss

$$p(x|\theta) \propto \prod_j \exp\left( - \frac{|f_j - f_\theta(x_j)|^2}{2\sigma_j^2} \right)$$

$$\Rightarrow \qquad \mathcal{L} \equiv -\log p(x|\theta) = \sum_j \frac{|f_j - f_\theta(x_j)|^2}{2\sigma_j^2}$$

# Training and uncertainties

## Learned scalar field $f_\theta(x) \approx f(x)$

- maximize parameter probability given $(f_j, \sigma_j)$

$$\theta = \text{argmax}\; p(\theta|x) = \text{argmax}\; \frac{p(x|\theta)\; p(\theta)}{p(x)}$$

$\to$ Gaussian likelihood loss

$$p(x|\theta) \propto \prod_j \exp\left( - \frac{|f_j - f_\theta(x_j)|^2}{2\sigma_j^2} \right)$$

$$\Rightarrow \qquad \mathcal{L} \equiv - \log p(x|\theta) = \sum_j \frac{|f_j - f_\theta(x_j)|^2}{2\sigma_j^2}$$

## Unknown uncertainties

- loss including normalization

$$\mathcal{L} = \frac{|f(x) - f_\theta(x)|^2}{2\sigma_\theta(x)^2} + \log \sigma_\theta(x) + \cdots$$

- if needed replace with Gaussian mixture model
- similar for Bayesian networks and evidential regression

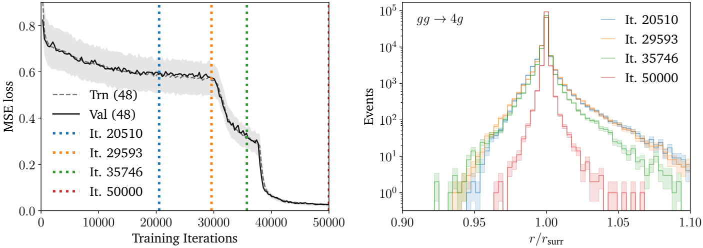
$\to$ Learning function and (systematic) uncertainty

# Symmetry and accuracy

## Amplitude ratio regression [Villadamigo, Frederix, TP, Vitos, Winterhalder]

- full vs. leading color ratio $r$ for $gg \to 4g$
- standard transformer training [not for MLP, GNN, L-GATr]

$\to$ Related to accuracy gain



# Symmetry and accuracy

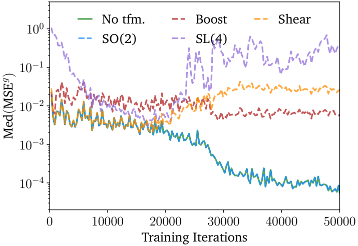
## Amplitude ratio regression [Villadamigo, Frederix, TP, Vitos, Winterhalder]

- full vs. leading color ratio $r$ for $gg \to 4g$
- standard transformer training [not for MLP, GNN, L-GATr]

$\to$ Related to accuracy gain

## Symmetries?

- evaluate MSE on transformed data
  wrong: 4D rotation SL(4), $y - z$-shear
  right: Lorentz boosts, SO(2) rotations
- initially learning general structures
  then penalizing wrong symmetries
  Lorentz symmetry stable
  SO(2) trivially good
- Symmetries help accuracy

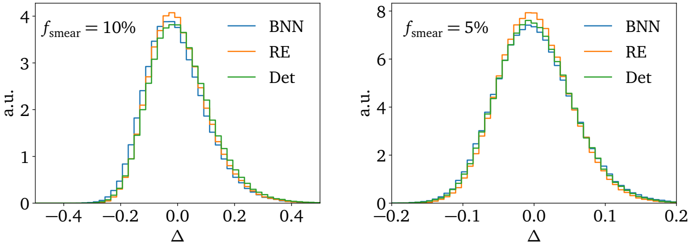


# Amplitudes with calibrated uncertainties

## Amplitude regression [Badger, Butter, Luchmann, Pitz, TP]

- loop amplitude $gg \to \gamma\gamma g(g)$ over
- systematics: artificial noise
- statistics plateau
- accuracy over phase space

$$\Delta(x) = \frac{A_{\text{NN}}(x) - A_{\text{true}}(x)}{A_{\text{true}}(x)}$$



# Amplitudes with calibrated uncertainties

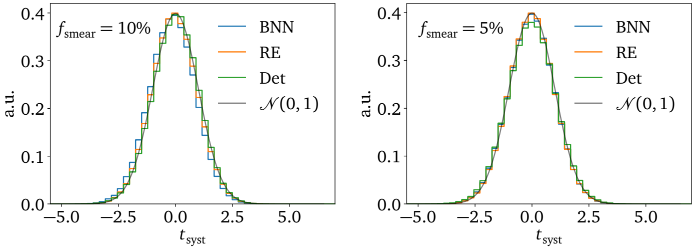
## Amplitude regression [Badger, Butter, Luchmann, Pitz, TP]

- loop amplitude $gg \to \gamma\gamma g(g)$ over
- systematics: artificial noise
- statistics plateau
- accuracy over phase space

$$\Delta(x) = \frac{A_{\text{NN}}(x) - A_{\text{true}}(x)}{A_{\text{true}}(x)}$$

- pull over phase space

$$t_{\text{syst}}(x) = \frac{A_{\text{NN}}(x) - A_{\text{true}}(x)}{\sigma_{\text{syst}}(x)}$$



# Amplitudes with calibrated uncertainties

## Amplitude regression [Badger, Butter, Luchmann, Pitz, TP]

- loop amplitude $gg \to \gamma\gamma g(g)$ over
- systematics: artificial noise
- statistics plateau
- accuracy over phase space

$$\Delta(x) = \frac{A_\text{NN}(x) - A_\text{true}(x)}{A_\text{true}(x)}$$

- pull over phase space

$$t_\text{syst}(x) = \frac{A_\text{NN}(x) - A_\text{true}(x)}{\sigma_\text{syst}(x)}$$

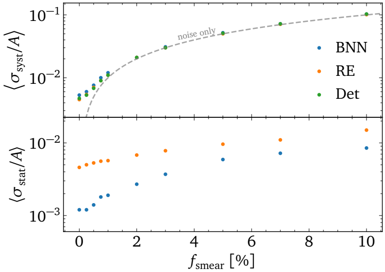
## Towards zero noise

- scaling

$$\sigma^2_\text{syst} - \sigma^2_\text{syst,0} \approx \sigma^2_\text{train}$$

- plateau $\langle \sigma_\text{syst}/A \rangle \sim 0.4\%$

$\to$ Limiting factor??



# Amplitudes with calibrated uncertainties

## Amplitude regression [Badger, Butter, Luchmann, Pitz, TP]

- loop amplitude $gg \to \gamma\gamma g(g)$ over
- systematics: artificial noise
- statistics plateau
- accuracy over phase space

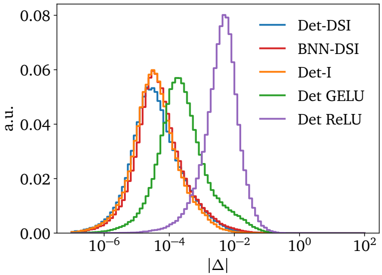
$$\Delta(x) = \frac{A_{\text{NN}}(x) - A_{\text{true}}(x)}{A_{\text{true}}(x)}$$

- pull over phase space

$$t_{\text{syst}}(x) = \frac{A_{\text{NN}}(x) - A_{\text{true}}(x)}{\sigma_{\text{syst}}(x)}$$

## Data pre-processing

- amplitude from invariants
- learn Minkowski metric
- Deep-sets-invariant network
  L-GATr transformer the same



# Amplitudes with calibrated uncertainties

## Amplitude regression [Badger, Butter, Luchmann, Pitz, TP]

- loop amplitude $gg \to \gamma\gamma g(g)$ over
- systematics: artificial noise
- statistics plateau
- accuracy over phase space

$$\Delta(x) = \frac{A_{\text{NN}}(x) - A_{\text{true}}(x)}{A_{\text{true}}(x)}$$

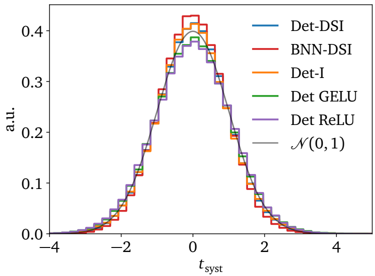
- pull over phase space

$$t_{\text{syst}}(x) = \frac{A_{\text{NN}}(x) - A_{\text{true}}(x)}{\sigma_{\text{syst}}(x)}$$

## Data pre-processing

- amplitude from invariants
- learn Minkowski metric
- Deep-sets-invariant network
  L-GATr transformer the same

$\to$ Calibrated systematics



# ATLAS calibration

## Energy calibration with uncertainties [ATLAS + Heimel, TP, Vogel]

- interpretable calorimeter phase space $x$
- learned calibration function

$$\mathcal{R}_{\text{NN}}(x) \pm \Delta\mathcal{R}_{\text{NN}}(x) \approx \frac{E^{\text{obs}}(x)}{E^{\text{dep}}(x)}$$

- trained on simulations, statistics neglibigle
- systematics: noise in data
  network expressivity
  data representation ...

# ATLAS calibration

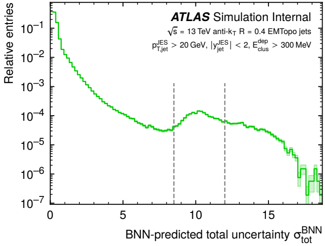
## Energy calibration with uncertainties [ATLAS + Heimel, TP, Vogel]

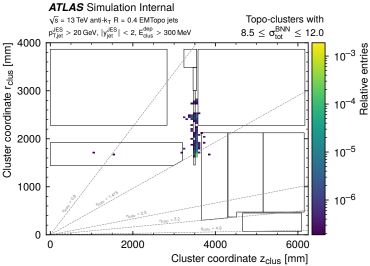
- interpretable calorimeter phase space $x$
- learned calibration function

$$\mathcal{R}_{\text{NN}}(x) \pm \Delta\mathcal{R}_{\text{NN}}(x) \approx \frac{E^{\text{obs}}(x)}{E^{\text{dep}}(x)}$$

- trained on simulations, statistics neglibigle
- systematics: noise in data
  network expressivity
  data representation ...

→ Understand (simulated) detector

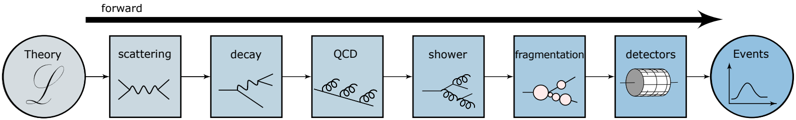


# Generative AI

## Simulations, MadNIS, calorimeters,...

- learn phase space density
  fast sampling Gaussian → phase space
  Bayesian generative network → uncertainties
- Variational Autoencoder
  → low-dimensional physics
- GAN [Butter, TP, Winterhalder]
  → generator trained by classifier
- Normalizing Flow [Bellagente, Haußmann, Luchmann, TP]
  → bijective mapping
- Diffusion [Butter, Hütsch, Palacios, TP, Sorrenson, Spinner]
  → ODE solving
- JetGPT, ViT [Favaro, Ore, Palacios, TP]
  → non-local transformers
- L-GATr / LLoCa [Brehmer, Breso, de Haan, TP, Qu, Spinner, Thaler]
  → Lorentz-covariant data representation



# Extrapolating ISR

## Universal QCD jet radiation [$Z + 1 \ldots 8$ jets]

- from $n$ to $n+1$ jets

$$R_{(n+1)/n} = \frac{\sigma_{n+1}}{\sigma_n} \qquad \text{and} \qquad P(n) = \frac{\sigma_n}{\sigma_{\text{tot}}} \qquad \text{with} \qquad \sigma_{\text{tot}} = \sum_{n=0}^{\infty} \sigma_n \; .$$

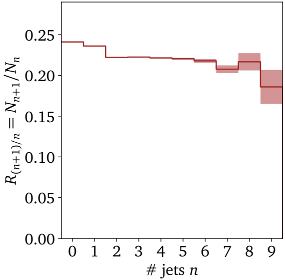
- large scale drop: Poisson scaling

$$R_{(n+1)/n} = \frac{\bar{n}}{n+1} \qquad \Leftrightarrow \qquad P(n) = \frac{\bar{n}^n e^{-\bar{n}}}{n!} \; .$$

- democratic scales: staircase scaling

$$R_{(n+1)/n} = e^{-b} = 1 - \tilde{\Delta}_g(Q^2) \equiv P(n+1|n)$$

$\to$ Universal pattern learnable?



# Extrapolating ISR

## Universal QCD jet radiation [$Z + 1 \ldots 8$ jets]

- democratic scales: staircase scaling

$$R_{(n+1)/n} = e^{-b} = 1 - \tilde{\Delta}_g(Q^2) \equiv P(n+1|n)$$

$\to$ Universal pattern learnable?

## Autoregressive transformer [Butter, Charton, Villadamigo, Ore, TP, Spinner]

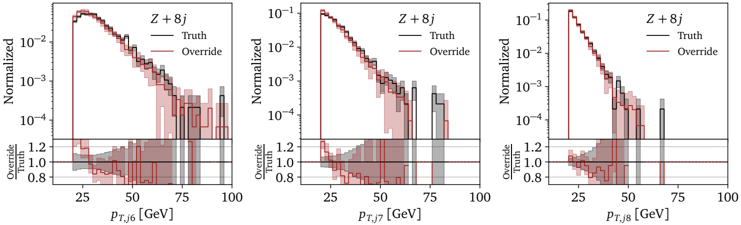
- factorized probability and loss function

$$p(x_i|x_{1:i-1}) = p_{\text{kin}}(x_i|x_{1:i-1})\; p_{\text{split}}(x_{1:i-1})$$

$$p(x_{1:n}) = \left[\prod_{i=1}^{n} p_{\text{kin}}(x_i|x_{1:i-1})\right] \left[\prod_{i=1}^{n} p_{\text{split}}(x_{1:i-1})\right] \left[1 - p_{\text{split}}(x_{1:n})\right] ,$$

- train up to 6 jets, generate 7 and 8

$\to$ full extrapolation with right latent representation



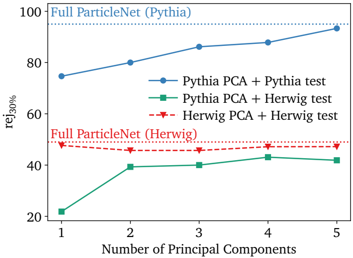
# Physics from latent representation

## Quarks vs gluons from trained ParticleNet [Vent, Winterhalder, TP]

- sensitive substructure variables

$$n_{\text{pf}} = \sum_i 1 \qquad w_{\text{pf}} = \frac{\sum_i p_{T,i} \Delta R_{i,\text{jet}}}{p_{T,\text{jet}}} \qquad p_T D = \frac{\sqrt{\sum_i p_{T,i}^2}}{\sum_i p_{T,i}} \qquad C_\beta = \frac{\sum_{i<j} p_{T,i} p_{T,j} (\Delta R_{ij})^\beta}{\left(\sum_i p_{T,i}\right)^2}$$

- related to max 5 principle components? [linear decorrelation]



# Physics from latent representation

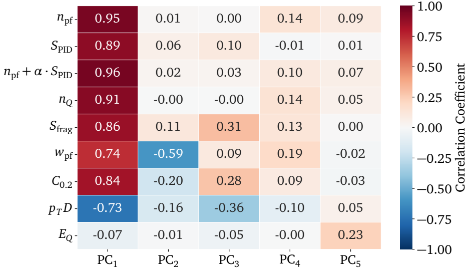
## Quarks vs gluons from trained ParticleNet [Vent, Winterhalder, TP]

- sensitive substructure variables

$$n_{\text{pf}} = \sum_i 1 \qquad w_{\text{pf}} = \frac{\sum_i p_{T,i} \Delta R_{i,\text{jet}}}{p_{T,\text{jet}}} \qquad p_T D = \frac{\sqrt{\sum_i p_{T,i}^2}}{\sum_i p_{T,i}} \qquad C_\beta = \frac{\sum_{i<j} p_{T,i} p_{T,j} (\Delta R_{ij})^\beta}{\left(\sum_i p_{T,i}\right)^2}$$

- related to max 5 principle components? [linear decorrelation]
- $\text{PC}_1$: constituent number and diversity

$$n_{\text{pf}} + \alpha \cdot S_{\text{PID}} \qquad \text{with} \qquad S_{\text{PID}} = -\sum_{\text{type}j} f_j \log f_j$$



| | $\text{PC}_1$ | $\text{PC}_2$ | $\text{PC}_3$ | $\text{PC}_4$ | $\text{PC}_5$ |
|---|---|---|---|---|---|
| $n_{\text{pf}}$ | 0.95 | 0.01 | 0.00 | 0.14 | 0.09 |
| $S_{\text{PID}}$ | 0.89 | 0.06 | 0.10 | -0.01 | 0.01 |
| $n_{\text{pf}} + \alpha \cdot S_{\text{PID}}$ | 0.96 | 0.02 | 0.03 | 0.10 | 0.07 |
| $n_Q$ | 0.91 | -0.00 | -0.00 | 0.14 | 0.05 |
| $S_{\text{frag}}$ | 0.86 | 0.11 | 0.31 | 0.13 | 0.00 |
| $w_{\text{pf}}$ | 0.74 | -0.59 | 0.09 | 0.19 | -0.02 |
| $C_{0.2}$ | 0.84 | -0.20 | 0.28 | 0.09 | -0.03 |
| $p_T D$ | -0.73 | -0.16 | -0.36 | -0.10 | 0.05 |
| $E_Q$ | -0.07 | -0.01 | -0.05 | -0.00 | 0.23 |

# Physics from latent representation

## Quarks vs gluons from trained ParticleNet [Vent, Winterhalder, TP]

- sensitive substructure variables

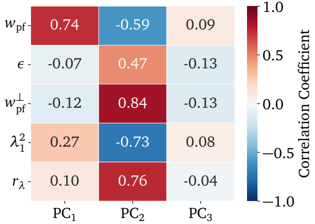
$$n_{\text{pf}} = \sum_i 1 \qquad w_{\text{pf}} = \frac{\sum_i p_{T,i} \Delta R_{i,\text{jet}}}{p_{T,\text{jet}}} \qquad p_T D = \frac{\sqrt{\sum_i p_{T,i}^2}}{\sum_i p_{T,i}} \qquad C_\beta = \frac{\sum_{i<j} p_{T,i} p_{T,j} (\Delta R_{ij})^\beta}{\left(\sum_i p_{T,i}\right)^2}$$

- related to max 5 principle components? [linear decorrelation]
- PC$_1$: constituent number and diversity

$$n_{\text{pf}} + \alpha \cdot S_{\text{PID}} \qquad \text{with} \qquad S_{\text{PID}} = -\sum_{\text{type}j} f_j \log f_j$$

- PC$_2$: radial energy profile

$$w_{\text{pf}}^\perp = \alpha \cdot n_{\text{pf}} - w_{\text{pf}} \qquad \text{and} \qquad r_\lambda = \frac{\lambda_{0.5}^1}{\lambda_1^2} \qquad \lambda_k^\beta = \sum_i z_i^\beta \Delta R^k$$



# Physics from latent representation

## Quarks vs gluons from trained ParticleNet [Vent, Winterhalder, TP]

- sensitive substructure variables

$$n_{\text{pf}} = \sum_i 1 \qquad w_{\text{pf}} = \frac{\sum_i p_{T,i} \Delta R_{i,\text{jet}}}{p_{T,\text{jet}}} \qquad p_T D = \frac{\sqrt{\sum_i p_{T,i}^2}}{\sum_i p_{T,i}} \qquad C_\beta = \frac{\sum_{i<j} p_{T,i} p_{T,j} (\Delta R_{ij})^\beta}{\left(\sum_i p_{T,i}\right)^2}$$

- related to max 5 principle components? [linear decorrelation]
- $\text{PC}_1$: constituent number and diversity

$$n_{\text{pf}} + \alpha \cdot S_{\text{PID}} \qquad \text{with} \qquad S_{\text{PID}} = -\sum_{\text{type}j} f_j \log f_j$$

- $\text{PC}_2$: radial energy profile

$$w_{\text{pf}}^\perp = \alpha \cdot n_{\text{pf}} - w_{\text{pf}} \qquad \text{and} \qquad r_\lambda = \frac{\lambda_{0.5}^1}{\lambda_1^2} \qquad \lambda_k^\beta = \sum_i z_i^\beta \Delta R^k$$

- $\text{PC}_3$: fragmentation and energy dispersion

$$S_{\text{frag}} = -\sum_i z_i \log z_i$$

| | $\text{PC}_1$ | $\text{PC}_2$ | $\text{PC}_3$ |
|---|---|---|---|
| $\log(p_T D)$ | -0.78 | -0.15 | -0.37 |
| $\log(p_T D)^\perp$ | -0.03 | -0.24 | -0.60 |
| $C_{0.1}$ | 0.83 | -0.08 | 0.33 |
| $C_{0.1}^\perp$ | 0.08 | 0.18 | 0.56 |
| $S_{\text{frag}}$ | 0.86 | 0.11 | 0.31 |
| $S_{\text{frag}}^\perp$ | 0.05 | 0.26 | 0.64 |
| $B^\perp$ | 0.04 | 0.23 | 0.65 |
| $A^\perp$ | 0.06 | 0.26 | 0.68 |

Correlation Coefficient (scale: −1.0, −0.5, 0.0, 0.5, 1.0)

# Physics from latent representation

## Quarks vs gluons from trained ParticleNet [Vent, Winterhalder, TP]

- sensitive substructure variables

$$n_{\text{pf}} = \sum_i 1 \quad w_{\text{pf}} = \frac{\sum_i p_{T,i} \Delta R_{i,\text{jet}}}{p_{T,\text{jet}}} \quad p_T D = \frac{\sqrt{\sum_i p_{T,i}^2}}{\sum_i p_{T,i}} \quad C_\beta = \frac{\sum_{i<j} p_{T,i} p_{T,j} (\Delta R_{ij})^\beta}{\left(\sum_i p_{T,i}\right)^2}$$

- related to max 5 principle components? [linear decorrelation]
- $\text{PC}_1$: constituent number and diversity

$$n_{\text{pf}} + \alpha \cdot S_{\text{PID}} \qquad \text{with} \qquad S_{\text{PID}} = -\sum_{\text{type}j} f_j \log f_j$$

- $\text{PC}_2$: radial energy profile

$$w_{\text{pf}}^\perp = \alpha \cdot n_{\text{pf}} - w_{\text{pf}} \qquad \text{and} \qquad r_\lambda = \frac{\lambda_{0.5}^1}{\lambda_1^2} \qquad \lambda_k^\beta = \sum_i z_i^\beta \Delta R^k$$

- $\text{PC}_3$: fragmentation and energy dispersion

$$S_{\text{frag}} = -\sum_i z_i \log z_i$$

- $\text{PC}_{4,5}$: charge information etc

$$E_Q = \frac{E_{\text{charged}}}{E_{\text{jet}}} \qquad \text{and} \qquad A^\perp = S_{\text{frag}} \frac{C_{0.1}}{C_{0.05}} - 0.03 \cdot n_{\text{pf}} + 1.95 w_{\text{pf}}^\perp$$

→ Latent distributions learn physics

# ParticleNet beyond PCA

## Disentangled latent classifier

- learning compressed, decorrelated representation

$$\mathcal{L} = \underbrace{\sum_{i=1}^{N} |x_i - \hat{x}_i|^2}_{\mathcal{L}_\text{reco}} + \underbrace{\sum_{i=1}^{N} \Big[ y_i \log \sigma(z_i) + (1 - y_i) \log(1 - \sigma(z_i)) \Big]}_{\mathcal{L}_\text{class}} + \underbrace{\sum_{j \neq k} \Big[ \text{Cov}(z_j, z_k) \Big]^2}_{\mathcal{L}_\text{disentangle}}$$

→ 5 latent dimensions plenty

| Latent Dim | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| AUC | 0.893(2) | 0.9001(4) | 0.9024(4) | 0.9034(2) |
| $\text{rej}_{30\%}$ | 72(3) | 77(3) | ´ 95(5) | 95(3) |
| $\Delta C$ | 1.8(3) | 0.93(5) | 1.0(16) | 0.9(15) |

# ParticleNet beyond PCA

## Disentangled latent classifier

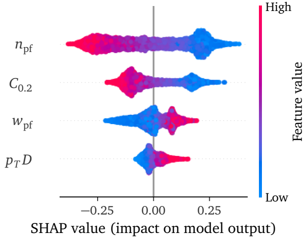
- learning compressed, decorrelated representation

$$\mathcal{L} = \underbrace{\sum_{i=1}^{N} |x_i - \hat{x}_i|^2}_{\mathcal{L}_\text{reco}} + \underbrace{\sum_{i=1}^{N} \Big[ y_i \log \sigma(z_i) + (1 - y_i) \log(1 - \sigma(z_i)) \Big]}_{\mathcal{L}_\text{class}} + \underbrace{\sum_{j \neq k} \Big[ \text{Cov}(z_j, z_k) \Big]^2}_{\mathcal{L}_\text{disentangle}}$$

$\to$ 5 latent dimensions plenty

## Shapley variables

- pretty pictures, with weird patterns [quark



# ParticleNet beyond PCA

## Disentangled latent classifier

- learning compressed, decorrelated representation

$$\mathcal{L} = \underbrace{\sum_{i=1}^{N} |x_i - \hat{x}_i|^2}_{\mathcal{L}_\text{reco}} + \underbrace{\sum_{i=1}^{N} \Big[ y_i \log \sigma(z_i) + (1 - y_i) \log(1 - \sigma(z_i)) \Big]}_{\mathcal{L}_\text{class}} + \underbrace{\sum_{j \neq k} \Big[ \text{Cov}(z_j, z_k) \Big]^2}_{\mathcal{L}_\text{disentangle}}$$

→ 5 latent dimensions plenty

## Shapley variables

- pretty pictures, with weird patterns [quark jets: low multiplicity and low girth]

→ Little insight from SHAP implementation

# ParticleNet beyond PCA

## Disentangled latent classifier

- learning compressed, decorrelated representation

$$\mathcal{L} = \underbrace{\sum_{i=1}^{N} |x_i - \hat{x}_i|^2}_{\mathcal{L}_{\text{reco}}} + \underbrace{\sum_{i=1}^{N} \Big[ y_i \log \sigma(z_i) + (1 - y_i) \log(1 - \sigma(z_i)) \Big]}_{\mathcal{L}_{\text{class}}} + \underbrace{\sum_{j \neq k} \Big[ \text{Cov}(z_j, z_k) \Big]^2}_{\mathcal{L}_{\text{disentangle}}}$$

→ 5 latent dimensions plenty

## Shapley variables

- pretty pictures, with weird patterns [quark jets: low multiplicity and low girth]

→ Little insight from SHAP implementation

## Symbolic Regression

- learn formula with given complexity
- classifier output not power series
- AUC and calibration double goal

$$p_{\text{quark}} = \tanh^3 \Big[ 0.55 \cdot C_{0.2} + 2 \left( -0.02 \cdot r_\lambda \cdot \left( C_{0.2} \cdot p_T D \cdot S_{\text{PID}} \cdot S_{\text{frag}} - 0.25 \right) + 1 \right)^3 \Big]$$

| observables | model | AUC | $\text{Rej}_{30\%}$ |
|---|---|---|---|
| $(n_{\text{pf}}, p_T D, C_{0.2}, r_\lambda, S_{\text{PID}}, S_{\text{frag}}, E_Q)$ | MLP | 0.872 | 66.87 |
| | PySR | 0.871 | 66.58 |

# ParticleNet beyond PCA

## Disentangled latent classifier

- learning compressed, decorrelated representation

$$\mathcal{L} = \underbrace{\sum_{i=1}^{N} |x_i - \hat{x}_i|^2}_{\mathcal{L}_\text{reco}} + \underbrace{\sum_{i=1}^{N} \Big[ y_i \log \sigma(z_i) + (1 - y_i) \log(1 - \sigma(z_i)) \Big]}_{\mathcal{L}_\text{class}} + \underbrace{\sum_{j \neq k} \Big[ \text{Cov}(z_j, z_k) \Big]^2}_{\mathcal{L}_\text{disentangle}}$$

→ 5 latent dimensions plenty

## Shapley variables

- pretty pictures, with weird patterns [quark jets: low multiplicity and low girth]

→ Little insight from SHAP implementation

## Symbolic Regression

- learn formula with given complexity
- classifier output not power series
- AUC and calibration double goal

$$p_\text{quark} = \tanh^3 \Big[ 0.55 \cdot C_{0.2} + 2 \left( -0.02 \cdot r_\lambda \cdot \left( C_{0.2} \cdot p_T D \cdot S_\text{PID} \cdot S_\text{frag} - 0.25 \right) + 1 \right)^3 \Big]$$

→ Formulas as physics regularizers? [Bahl, Fuchs, Menem, TP]

# ML for LHC Theory

## ML is particle physics method development

1 just another numerical tool for a numerical field

2 completely transformative new language

- driven by (money from) industry and medical research
- particle physics should be leading scientific AI
- improve established tools
  develop new tools for established tasks
  transform through new ideas

→ Complexity becoming our friend

arXiv:2211.01421v2 [hep-ph] 17 Mar 2024

Modern Machine Learning for LHC Physicists

Tilman Plehn[a,*], Anja Butter[a,b], Barry Dillon[a],
Theo Heimel[a], Claudius Krause[c], and Ramon Winterhalder[d]

[a] Institut für Theoretische Physik, Universität Heidelberg, Germany
[b] LPNHE, Sorbonne Université, Université Paris Cité, CNRS/IN2P3, Paris, France
[c] HEPHY, Austrian Academy of Sciences. Vienna, Austria
[d] CP3, Université catholique de Louvain, Louvain-la-Neuve, Belgium

March 19, 2024

**Abstract**

Modern machine learning is transforming particle physics fast, bullying its way into our numerical tool box. For young researchers it is crucial to stay on top of this development, which means applying cutting-edge methods and tools to the full range of LHC physics problems. These lecture notes lead students with basic knowledge of particle physics and significant enthusiasm for machine learning to relevant applications. They start with an LHC-specific motivation and a non-standard introduction to neural networks and then cover classification, unsupervised classification, generative networks, and inverse problems. Two themes defining much of the discussion are well-defined loss functions and uncertainty-aware networks. As part of the applications, the notes include some aspects of theoretical LHC physics. All examples are chosen from particle physics publications of the last few years.[1]